DOCUMENT RESUME

ED 058 317 TM 001 033

AUTHOR Urry, Vern W.
TITLE Approximation Methods for the Item Parameters of Mental Test Models.
INSTITUTION Washington Univ., Seattle. Bureau of Testing.
PUB DATE Dec 71
NOTE 27p.

EDRS PRICE MF-$0.65 HC-$3.29
DESCRIPTORS Correlation; Discriminant Analysis; Guessing (Tests); *Item Analysis; *Mental Tests; Models; *Research Methodology; Scores; *Statistical Analysis; *Test Construction

ABSTRACT

Equations are derived to enable the graphic approximation of the item parameters of the stochastic mental test models, i.e., the generalized normal ogive and logistic models. The item parameters for the models are discriminatory power, difficulty, and probability of chance success. Suggested uses for the approximations were to provide a basis for screening items for tailored testing, to enable a determination as to the appropriateness of a set of items for tailored testing, and to provide starting values for parameter estimation in maximum likelihood procedures. The conditions and assumptions necessary for an effective application of the method were delineated. Recent empirical results which bear on the properties of the approximations were examined. An investigation was suggested to evaluate a further possible use of the approximations--that of their direct applicability in tailored testing procedures. (Author/DLG)

Bureau of Testing
University of Washington—Seattle

Bureau of Testing

University of Washington

December 1971

Approximation Methods for the Item Parameters of Mental Test Models

Vern W. Urry

Equations were derived to enable the graphic approximation of the item parameters of the stochastic mental test models, i.e., the generalized normal ogive and logistic models. The item parameters for the models are discriminatory power $(\underline{a}_i)$, difficulty $(\underline{b}_i)$, and probability of chance success $(\underline{c}_i)$. In brief, the probability of chance success on the item $(\underline{c}_i)$ can be approximated through visual inspection of the left-hand (lower) asymptote of the proportion passing the item plotted against the total test score minus the particular item. Thereafter, a graph appropriate to the approximated $\underline{c}_i$ can be consulted to convert an ordinary item-total test point-biserial correlation and proportion passing the item into approximations of item discriminatory power $(\underline{a}_i)$ and item difficulty $(\underline{b}_i)$. Suggested uses for the approximations were to provide a basis for screening items for tailored testing, to enable a determination as to the appropriateness of a set of items for tailored testing, and to provide starting values for parameter estimation in maximum likelihood procedures. The conditions and assumptions necessary for an effective application of the method were delineated. Recent empirical results which bear on the properties of the approximations were examined. An investigation was suggested to evaluate a further possible use of the approximations—that of their direct applicability in tailored testing procedures.

Bureau of Testing Project: 0871-202

Approximation Methods for the Item Parameters of Mental Test Models

Vern W. Urry

In item analytic procedures, two indices, item discrimination and item difficulty, are conventionally computed. The index of item discrimination is usually a biserial or point-biserial correlation between the item and the total test score. If the sample size is sufficiently large and the test is of sufficient length and homogeneity, this index bears a close relationship to the correlation between the item and the latent ability measured by the test. The latter correlation is the parameter of item discrimination of conventional interest. The correlation between the item and total test score is test dependent; whereas, the correlation between an item and latent ability is generalizable. The index of item difficulty usually computed is the proportion passing the item. In this instance, the conventional index is also an estimate of an item parameter, the population proportion passing the item.

Under certain conditions and assumptions, these estimates of the parameters of conventional interest can also be used to obtain approximations to the item parameters of stochastic mental test models. The stochastic models we are concerned with are the generalized logistic and normal ogive models.[1] The model parameters we wish to approximate are item discriminatory

[1]The generalized normal ogive model and the generalized logistic model are considered interchangeable in the present report. Indeed the generalized logistic model may be viewed as a mathematically convenient form of the generalized normal ogive model. The generalized model in both cases subsumes the instances where guessing is or is not effective and the item discriminatory powers are or are not constant. The interrelationships among the four logistic models have been delineated elsewhere (Urry, 1970). The generalized logistic model is the Birnbaum 3-parameter model. Its normal ogive counterpart, strictly speaking, is the subject of the present report and has been alluded to earlier by Lord (1968b, p. 384).

power $(\underline{a}_i)$, item difficulty $(\underline{b}_i)$, and probability of chance success on the item $(\underline{c}_i)$.

The stochastic mental test models enable applications which have captured the imagination of many researchers. An example is tailored testing, where certain requirements need to be placed on the model parameters to ensure an advantage of the tailored test over the usual paper and pencil test. In view of these requirements, items need to be screened for appropriateness. For instance, a sizeable number of appropriate items is necessary to constitute a bank from which sequential selections are made to form an individualized or tailored test. A repetitive cycle is involved. An item is selected, the item is responded to by the examinee, and the examinee's ability is (re-) estimated. The cycles terminate when the information regarding an examinee's ability is considered sufficient. In the establishment of an item bank, the separate items are deemed appropriate when they conform to a configuration of item parameters. The configuration found most effective in maintaining the superiority of tailored tests over paper and pencil tests (Urry, 1970; 1971) consists of items with item discriminatory powers $(\underline{a}_i)$ greater than 0.8, enjoying a rectangular distribution of item difficulties $(\underline{b}_i)$, say, from -2.0 to 2.0, and with probabilities of chance success $(\underline{c}_i)$ less than or equal to .25. Given such a configuration, tailored tests of 10 to 15 items may have a validity equal to that of a paper and pencil test several times its length.

Approximations of the model parameters for a set of items may be used to assess the efficacy of the set for the purpose of tailored testing. Normally, inappropriate items would then be rejected and the residual set

re-evaluated in view of the required configuration. Any deficiencies would then require the use of additional items. Maximum likelihood estimates of the model parameters for the final item bank would then be obtained through the iterative method suggested by Birnbaum (1968) and Lord (1968a). Since conventional procedures are, by far, the more economical, their use in the initial screening of items effects considerable savings. The number of items for which maximum likelihood estimates of the parameters are obtained is appropriately reduced, and the initial approximations can be used as the starting values for the parameter estimates required by the maximum likelihood equations. The convergence of the iterative method is hastened through the use of good initial values for the parameter estimates.

We have already noted that the effect of guessing is made explicit in the stochastic models. In conventional procedures it is largely neglected. Since the presence of guessing has a systematic effect on the parameters conventionally estimated, the probability of chance success on the item (c_i) must be approximated first in the application of the proposed method. In rough work, one may take as the approximate value the reciprocal of the number of item alternatives. More precisely, however, one could approximate the value from the left-hand asymptote of the proportion passing the item regressed on the total test score minus the particular item. Lord (1968a) has discussed the procedure in connection with maximum likelihood estimation of the parameters of the Birnbaum 3-parameter logistic model.

The Equations

We are concerned with two cases. The first case represents the situation where the items are in a free-response form. The second case is the instance where guessing is effective or when the items are of a multiple-choice type.

Free-response case

When items are in a free-response mode, i.e., the c_i equal zero or the number of alternatives is infinite, the approximation method can be rather direct. For a test of sufficient length and homogeneity, the item-total test biserial correlation based on a large sample provides a close approximation to $\rho_{I\theta}$, the biserial correlation between an item and the latent ability (θ). In consonance with the definition of the biserial correlation, the subscript I indexes the normally (and continuously) distributed variable assumed basic to the binary item. Some authors, most notably Brogden (1971), refer to this hypothetical variable as the item ghost. In contradistinction, the subscript i indexes the manifest binary or dichotomous item.

Item discriminatory power (a_i) may be approximated through

$$(1) \qquad a_i \equiv \frac{\rho_{I\theta}}{\sqrt{1 - \rho_{I\theta}^2}} .$$

For convenience of reference, several values of a_i are presented for corresponding values of $\rho_{I\theta}$ in Table 1. We observe through equation (1) that a_i is zero when $\rho_{I\theta}$ is zero and infinite when $\rho_{I\theta}$ is unity. Negative values of a_i or $\rho_{I\theta}$ are not of interest.

Now the proportion passing a free-response item (P_I) is given by the model as

Table 1

Values of Item Discriminatory Power (a_i) Corresponding to the Biserial Correlation Coefficient ($\rho_{I\theta}$)

a_i	$\rho_{I\theta}$
3.0	.949
2.0	.894
1.6	.848
1.4	.814
1.2	.768
1.0	.707
.8	.625

(2) $$P_I = (2\pi)^{-\frac{1}{2}} \int_{\gamma_i}^{\infty} \exp\left[-\frac{t^2}{2}\right] dt$$

where γ_i is the point of cut on the continuous and normal distribution underlying the binary item. In words, we consult a standard table of the normal distribution. The baseline value is sought which isolates the upper area corresponding to the estimated proportion passing the item. The baseline value is the approximate value of γ_i. Since

(3) $$\gamma_i = \rho_{I\theta} \underline{b}_i \quad ,$$

item difficulty $(\underline{b}_i)$ can be approximated through

(4) $$\underline{b}_i = \frac{\gamma_i}{\rho_{I\theta}} \quad .$$

While the above approach is both available and direct, few test items in current use are of the free-response type. Also the above approach does not generalize to the case where guessing is effective. For that case we will use large sample estimates of the point-biserial correlation and the proportion passing the item to obtain approximations to the model parameters. The use of the same sample estimates is possible for the free-response case. In fact, the equations may be illustrated as a special case subsumed under the general case where guessing is effective.

In the previous approach, approximations to the model parameters were obtained computationally; whereas, in the present approach, the model parameters are approximated graphically. In the graphic approach, the

equations are used to obtain exact values of the conventional parameters assuming the model parameters to be known. The results may then be graphed to allow the interpolation of the model parameters given large sample estimates of the conventional parameters.

The subsequent equations are provided for the free-response case to provide continuity in the present approach. We will need the relation

$$(5) \qquad \rho_{I\theta} = \frac{\underline{a}_i}{\sqrt{1 + \underline{a}_i^2}}$$

which is readily derived from equation (1). Now the variance of a binary item, subscripted in lower case, is

$$(6) \qquad \sigma_i^2 = P_I Q_I$$

where P_I is evaluated through equations (2) and (3). The proportion failing an item, Q_I, is obtained from

$$(7) \qquad Q_I = 1 - P_I \quad .$$

The covariance of a binary item and underlying ability is given by Brogden (1971) as

$$(8) \qquad \sigma_{i\theta} = \rho_{I\theta}\, \phi(\gamma_i)$$

where $\rho_{I\theta}$ is given by equation (5) and

$$(9) \qquad \phi(\gamma_i) = (2\pi)^{-\frac{1}{2}} \exp\left[-\frac{\gamma_i^2}{2} \right]$$

is the height of the ordinate at the point of cut on the continuous distribution assumed basic to the binary item. Since underlying ability (θ) is assumed $N(0,1)$, the point-biserial correlation between a binary item and underlying ability is

(10) $$\rho_{i\theta} = \frac{\sigma_{i\theta}}{\sigma_i} = \frac{\rho_{I\theta}\,\phi(\gamma_i)}{\sqrt{P_I Q_I}} \quad .$$

Thus for assumed values of $\underline{a}_i$ and $\underline{b}_i$ ($\underline{c}_i$ is null) the population values of P_I and $\rho_{i\theta}$ can be calculated. Given $\underline{a}_i$, $\rho_{I\theta}$ can be obtained through equation (5). Since $\underline{b}_i$ is known, the point of cut, γ_i , on the continuous distribution assumed basic to the binary item is given by equation (3). The proportion passing the item, P_I , is evaluated by equation (2). After employing equation (7), the variance of the binary item, σ_i^2 , is obtained through equation (6). The height of the ordinate at the point of cut, $\phi(\gamma_i)$, is given by equation (9). At this point, we have P_I and the necessary values to solve for $\rho_{I\theta}$ by means of equation (10). This feature will allow graphic approximations to the model parameters.

Multiple-choice case

When items are in multiple-choice form, the probabilities of chance success ($\underline{c}_i$) are usually not zero. In this instance, the population proportion passing an item and point-biserial correlation between a binary item and underlying ability are designated by $P_{I'}$ and $\rho_{i'\theta}$, respectively. The primes denote the presence of guessing. Success due to guessing inflates $P_{I'}$ and attenuates $\rho_{i'\theta}$. The magnitude of $\underline{c}_i$ is directly related to the degree of this distortion.

When items are affected by guessing, the proportion passing an item is given by

$$P_{I'} = \underline{c}_i + (1 - \underline{c}_i)\ P_I \tag{11}$$

where P_I is obtained through equation (2). Since

$$Q_{I'} = 1 - P_{I'} \quad , \tag{12}$$

the variance of the binary item is merely

$$\sigma_{i'}^2 = P_{I'} Q_{I'} \quad . \tag{13}$$

The covariance between a binary item affected by guessing and the latent ability was derived by Brogden (1971) as

$$\sigma_{i'\theta} = (1 - \underline{c}_i)\ \rho_{I\theta}\ \phi(\gamma_i) \quad . \tag{14}$$

where $\phi(\gamma_i)$ is given by equation (9). Therefore, the point-biserial correlation is

$$\rho_{i'\theta} = \frac{\sigma_{i'\theta}}{\sigma_{i'}} = \frac{(1 - \underline{c}_i)\ \rho_{I\theta}\ \phi(\gamma_i)}{\sqrt{P_{I'} Q_{I'}}} \quad . \tag{15}$$

Now if $\underline{c}_i$ is zero, a review of equations (11) through (15) will indicate that the results of equations (2) and (10) are respectively identical to the results of equations (11) and (15). When the population proportion passing an item and the point-biserial correlation are used in this manner, the free response case can be subsumed under the present case.

For purposes of illustration, we will employ equations (2), (3), (5), (9), and (11) through (15). Let us assume $\underline{a}_i$, $\underline{b}_i$, and $\underline{c}_i$ equal to 2.00, 1.50, and .20, respectively, for a given item i. Through equation (5) (or by obtaining the value directly from Table 1), $\rho_{I\theta}$ is found to be .894. The point of cut (γ_i) on the assumed underlying distribution of the item would be, through equation (3), 1.341. By equation (2), we note that P_I is obtained by getting the integral above γ_i in the normal distribution. Most readily, this is obtained from standard tables of the normal distribution. In our particular case, P_I is .090. Through equation (11), $P_{I'}$ is .20 plus .80 times .090 or .272. $Q_{I'}$, through equation (12), is .728. The height of the ordinate at γ_i, $\phi(\gamma_i)$, is given by equation (9) as .162. By direct substitution in equation (15) we have

$$\rho_{i'\theta} = \frac{(.80)(.894)(.162)}{\sqrt{(.272)(.728)}} = .260 \quad .$$

Given $\underline{c}_i$, $\rho_{i'\theta}$ and $P_{I'}$ can be calculated pairwise for any values of $\underline{a}_i$ and $\underline{b}_i$.

Method

Since $\rho_{i'\theta}$ and $P_{I'}$ can be calculated for systematic levels of $\underline{a}$ and $\underline{b}$, this will allow the establishment of a grid system for the model parameters $\underline{a}$ and $\underline{b}$ ($\underline{c}$ is held constant) on the coordinates determined by the conventionally estimated parameters, $P_{I'}$ and $\rho_{i'\theta}$. The curves obtained on the coordinate system when $\underline{a}$ or $\underline{b}$ is fixed define the grid system for the model parameters.

The systematic levels of $\underline{a}$ used in the present report are given in Table 1 where immediate translation to $\rho_{I\theta}$ is possible. The levels were 3.0, 2.0, 1.6, 1.4, 1.2, 1.0, and .8. No value less than .8 was used since item discriminatory powers of that magnitude are poorly suited to the purpose of tailored testing. The systematic levels of $\underline{b}$ were 2.0, 1.8,....-2.0. The range of values is also compatible with the objectives of tailored testing. In the present report the fixed levels of $\underline{c}$ were chosen at .00, .20, .25, .33, and .50. As pointed out earlier, these roughly correspond to free-response, 5-, 4-, 3- alternative and true-false item types.

Computer programs have been developed to provide the necessary data and to graph the results on a Calcomp Plotter. One graph is plotted for each specified level of $\underline{c}$. We discussed earlier a method of approximating the $\underline{c}_i$ from the left-hand asymptote of the proportion passing an item regressed on the adjusted total test score. If the method is applied, graphs of the appropriate probabilities of chance success should be used to enhance the accuracy of approximation.

Results

The generated graphs are presented as Charts 1 through 5, each corresponding to a specified level of $\underline{c}$, probability of chance success. The charts consist of the mapping of the grid system of the model parameters, $\underline{a}$ and $\underline{b}$, into a coordinate system where the population point-biserial correlation ($\rho_{i\theta}$ or $\rho_{i'\theta}$) is the ordinate and the population proportion passing an item (P_I or $P_{I'}$) is the abscissa. In each chart, the numeric values centrally and vertically arranged on the grid system denote the levels of $\underline{a}$, item discriminatory power. Located at the base of the grid system

are the numeric values for the levels of $\underline{b}$, item difficulty. From the charts, it may be noted that item difficulty ($\underline{b}$) is inversely related to the conventional measure of item difficulty, the proportion passing an item.

Displayed in Chart 1 is the free-response case where $\underline{c}$ equals zero. The grid system is symmetric about $\underline{b}$ equal to zero. As $\underline{b}$ departs from zero, the point-biserial correlation ($\rho_{i\theta}$) decreases even though the biserial correlation ($\rho_{I\theta}$) remains constant. For example, the curve where $\underline{a}$ equals 3.0 implies a constant biserial correlation of .949. Where the proportion passing an item is .03 and .97, the corresponding point-biserial correlations would both be .37. For $\underline{a}$ equal to 3.0, the maximum point-biserial correlation would be .76. The maximum point-biserial correlation occurs at the proportion passing an item of .5 ($\gamma = \underline{b} = 0.0$). When the biserial correlation is unity, we can see through equation (1) that $\underline{a}$ is then infinite. The maximum value for the point-biserial correlation through equation (10) would be $\frac{\phi(\gamma = 0.0)}{\sqrt{(.5)(.5)}}$ or .80--twice the height of the ordinate in the normal distribution at a baseline value of zero.

When the probability of chance success is not zero, the symmetry observed in Chart 1 no longer obtains. This may be seen in Chart 2, where the grid system is provided for $\underline{c}$ equal to .20. It may be recalled that this level of probability of chance success approximates the case of items with five alternatives. Noticeably, the grid system of the model parameters is displaced to the right. The displacement reflects the inflation of the proportion passing an item relative to the free-response case. Again relative to the results depicted in Chart 1, the point-biserial correlations are attenuated through error due to guessing. The attenuation is apparent

CHART 1
RELATIONSHIP BETWEEN CONVENTIONAL AND NORMAL OGIVE ITEM PARAMETERS
WHEN THE PROBABILITY OF CHANCE SUCCESS(C) EQUALS .00

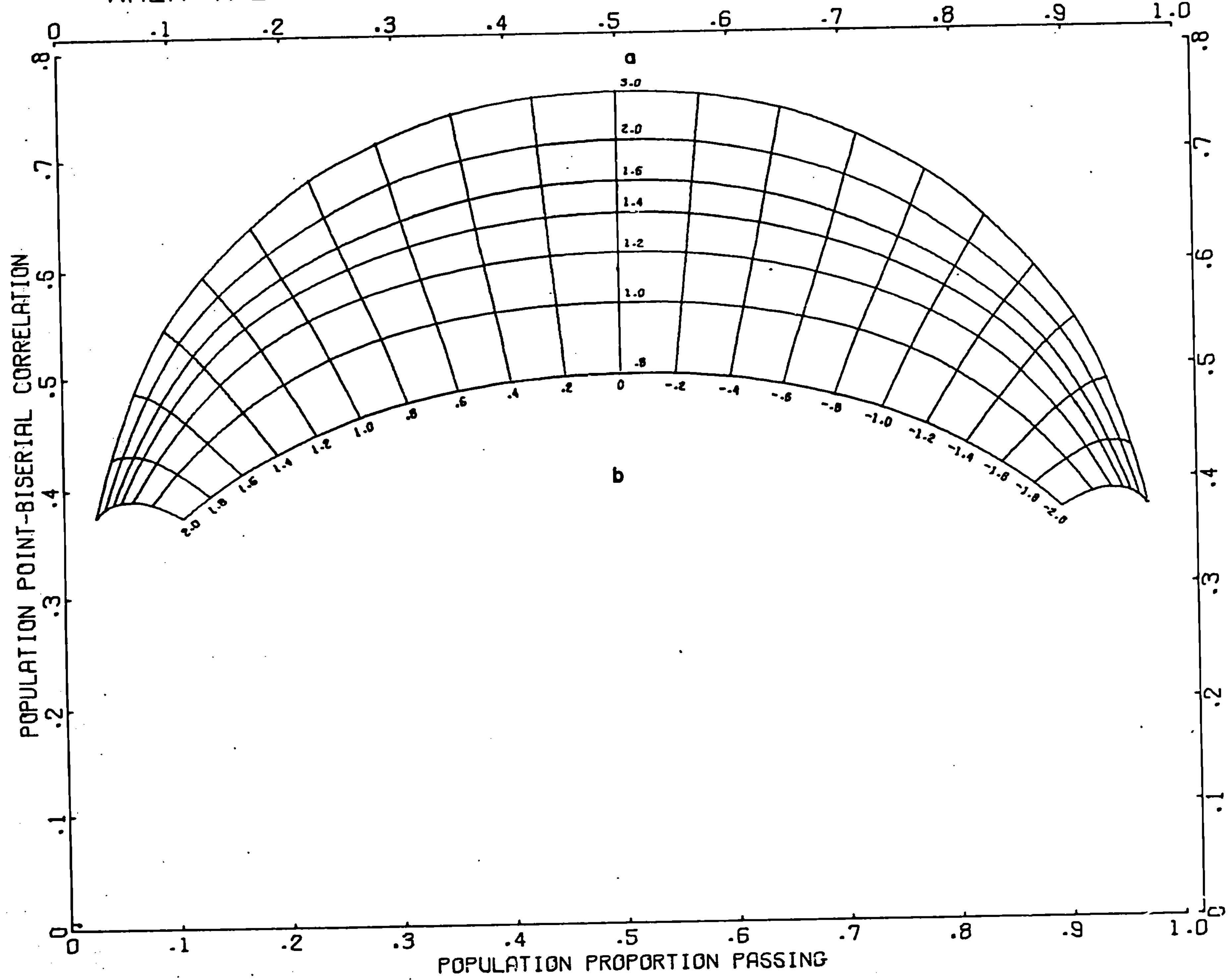
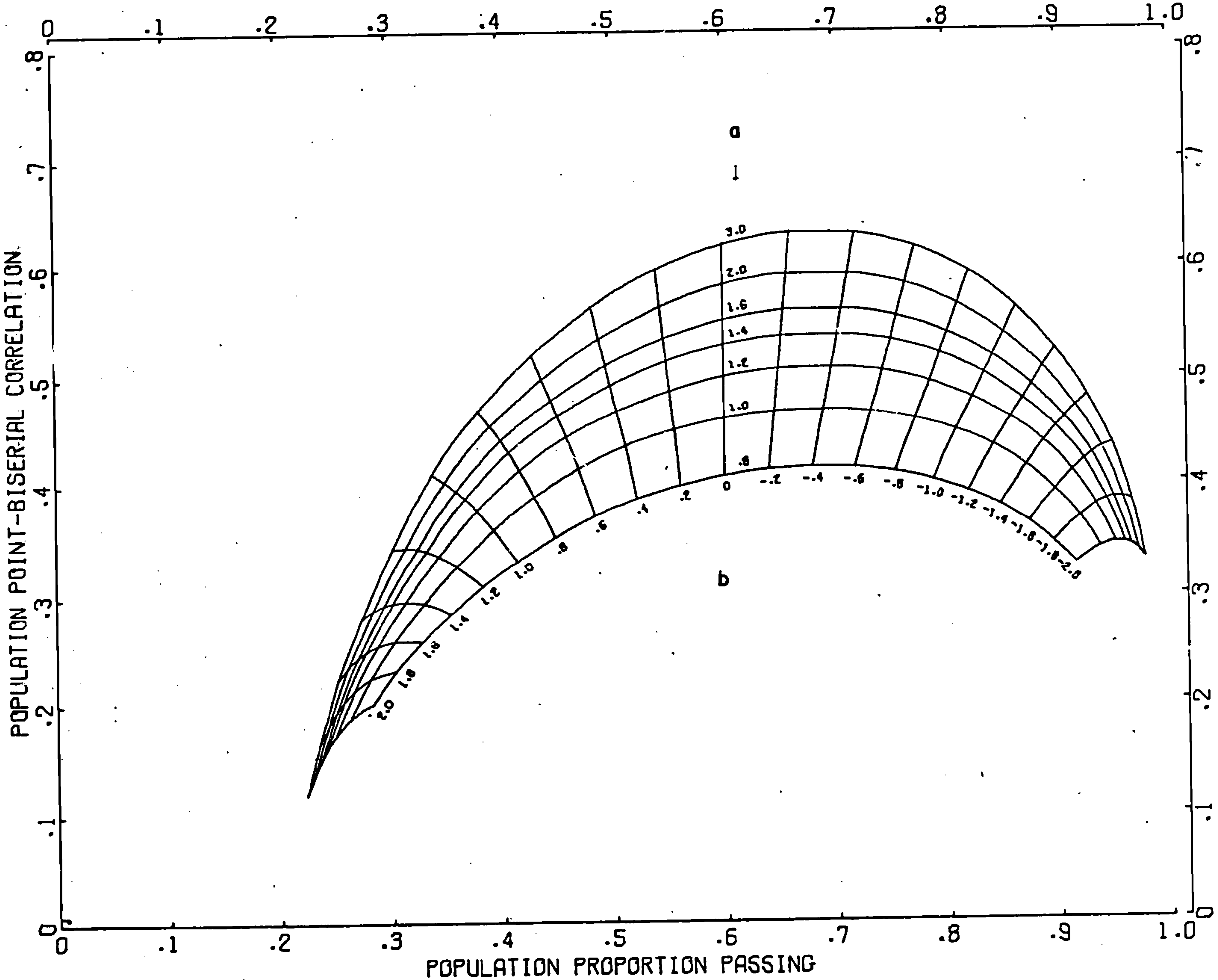


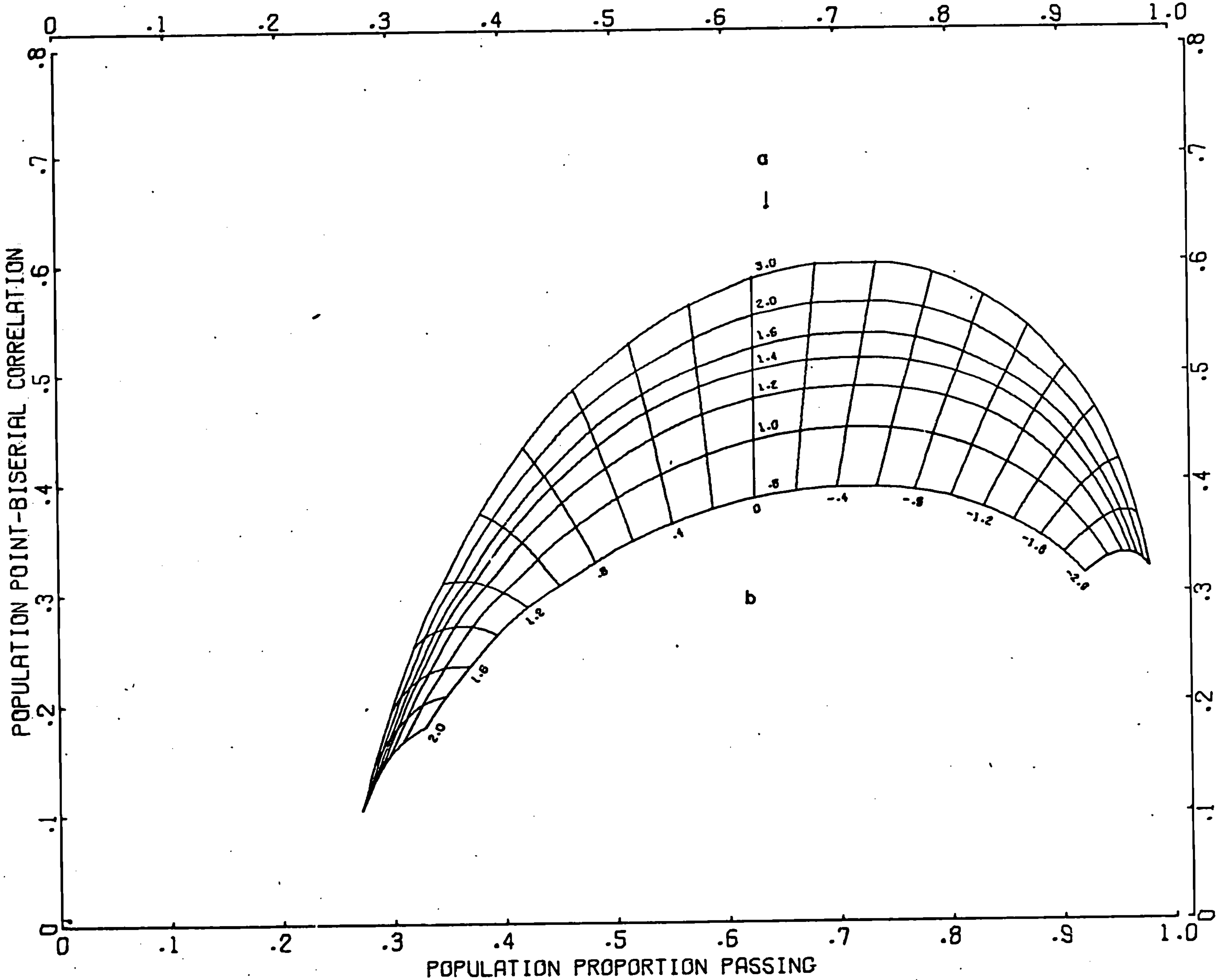
CHART 2
RELATIONSHIP BETWEEN CONVENTIONAL AND NORMAL OGIVE ITEM PARAMETERS
WHEN THE PROBABILITY OF CHANCE SUCCESS(C) EQUALS .20



in the general lowering of the grid system. Notice that the attenuation is more severe for items of greater difficulty, i.e., where b is positive or the proportion passing is low. If a is held constant at 3.0, the point-biserial correlation ranges in value from .12, at b equal to 2.0 or a proportion passing of .22, to .63, at b equal to -.30 or a proportion passing of .68. At b equal to -2.0 or a proportion passing of .98, the point-biserial correlation is .33. Even though the model parameter of item discriminatory power is held constant, there is still a range of values possible for the conventional measure of item discrimination, the point-biserial correlation.

The results obtained when c equals .25 are portrayed in Chart 3. In practice this case is approximated when 4-alternative items are used. The distortion due to the probability of chance success--the inflation of the probability of passing an item and the attenuation of the point-biserial correlation--becomes more pronounced. For example, if a is 3.0 and b is 2.0, the corresponding proportions passing an item and point-biserial correlation would be .27 and .11. Similarly, for a equal to 3.0 and b equal to -2.0, the proportion passing an item and the point-biserial correlation would be .98 and .32. The maximum value for the point-biserial correlation at a equal to 3.0 is .60 and occurs when b equals -.47 or the proportion passing is .76. If a for a given item is 3.0, that item is highly desirable because of its high level of discriminatory power; however, an item with a point-biserial correlation as low as .11, for instance, might under extant practices be rejected on the basis of what could be an erroneous conclusion, i.e., the item is not sufficiently discriminating. Hopefully,

CHART 3
RELATIONSHIP BETWEEN CONVENTIONAL AND NORMAL OGIVE ITEM PARAMETERS
WHEN THE PROBABILITY OF CHANCE SUCCESS(C) EQUALS .25



the point-biserial correlation would be examined with knowledge of the proportion passing the item as well as the degree of effectiveness of guessing.

In Chart 4, the data are graphed for c equal to .33 or roughly the 3-alternative item case. The distortion noted above becomes progressively more pronounced as the probability of chance success increases. The area encompassed by the grid system is simultaneously reduced. Again if a is held constant at 3.0, a b of 2.0 or a proportion passing of .35 provides a point-biserial correlation of .09. At the same level of item discriminatory power, the maximum point-biserial correlation of .56 occurs at b equal to -.50 or a proportion passing of .78. When b is -2.0 and a is 3.0, the proportion passing is .98 and the point-biserial correlation is .30. For purposes of illustration, a has been held at a constant value of 3.0. Analogous interpretations apply for any fixed value of item discriminatory power. The point is that when item discriminatory power is fixed the conventional measure of reputedly the same characteristic still has considerable variability.

For results paralleling the true-false item case, Chart 5 may be consulted. This is the case where the probability of chance success on the item equals .50. The distortion noted earlier becomes more severe. The proportion passing becomes more inflated and the degree of attenuation is more marked. The area encompassed by the grid system is again further diminished. To the extent that errors in the estimation of the proportion passing an item and the point-biserial correlation are dependent on sample size, the approximations of a and b for fixed sample size would become progressively worse as c increases. Again, if a is fixed at 3.0: a b of 2.0 provides a

CHART 4
RELATIONSHIP BETWEEN CONVENTIONAL AND NORMAL OGIVE ITEM PARAMETERS
WHEN THE PROBABILITY OF CHANCE SUCCESS(C) EQUALS .33

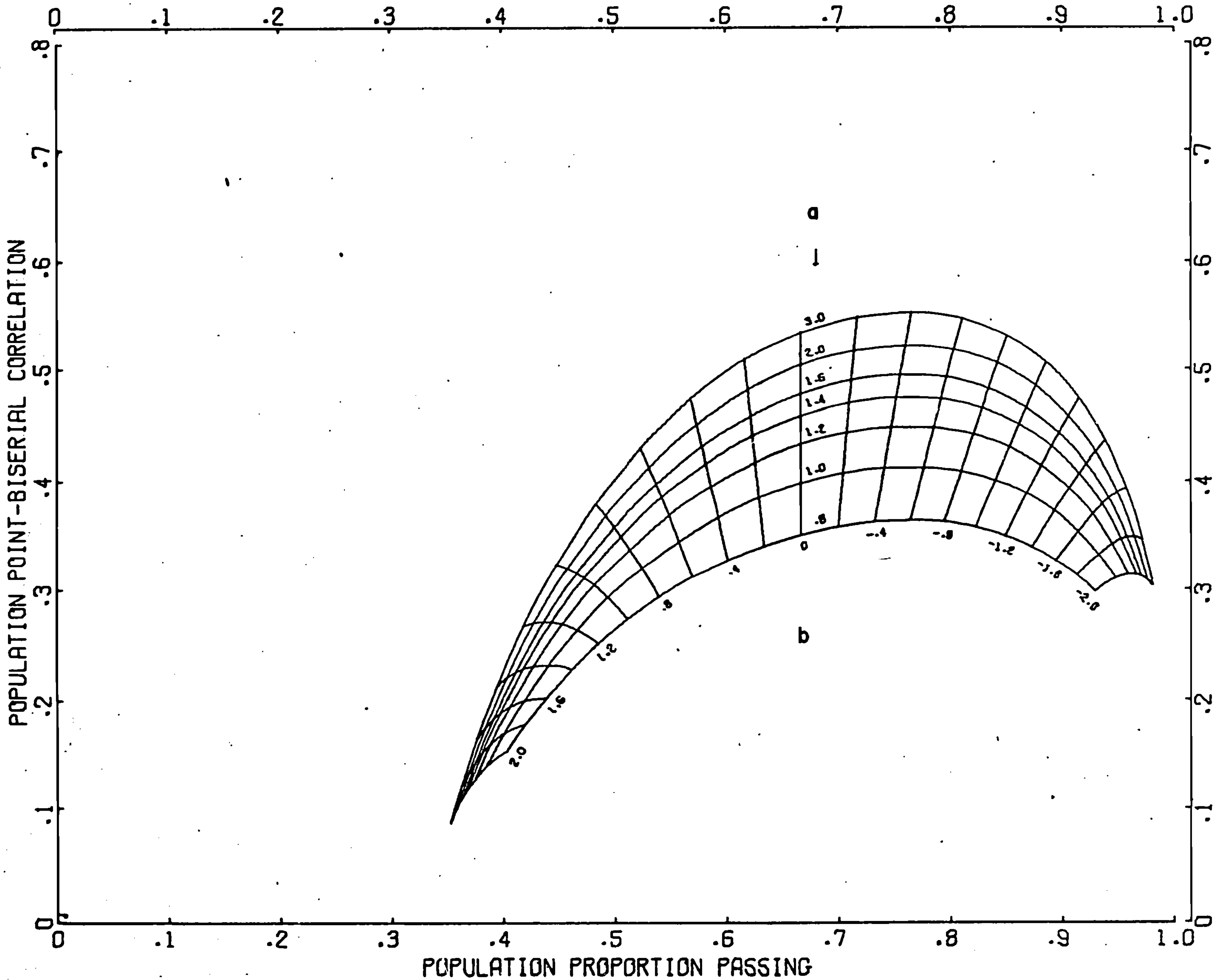
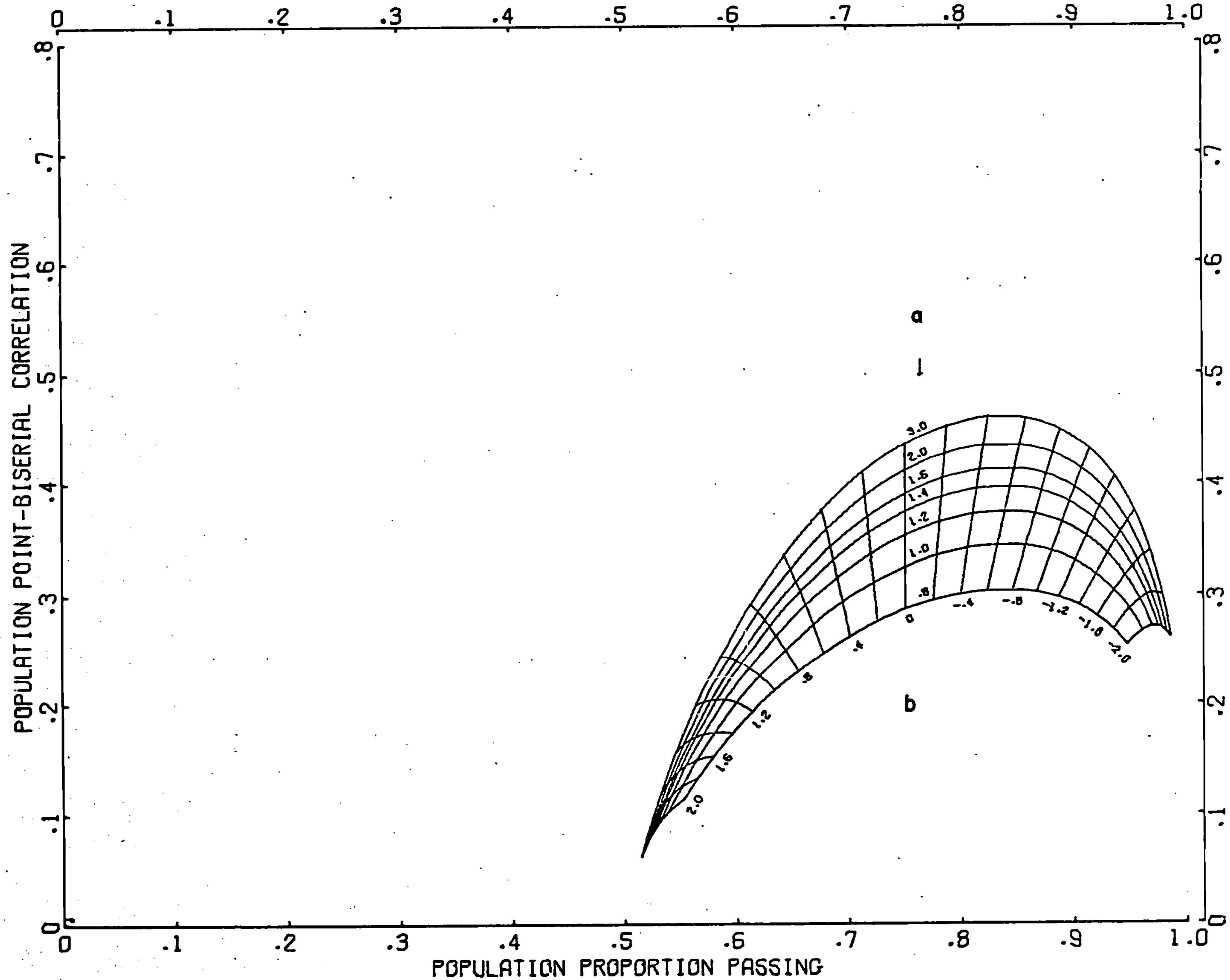


CHART 5

RELATIONSHIP BETWEEN CONVENTIONAL AND NORMAL OGIVE ITEM PARAMETERS WHEN THE PROBABILITY OF CHANCE SUCCESS(C) EQUALS .50



proportion passing of .50 and a point-biserial correlation of .06; a maximum for the point-biserial correlation, .46, is obtained at a b of -.55 where the proportion passing is .85; and a b of -2.0 corresponds to a proportion passing of .99 and a point-biserial correlation of .26. Item discriminatory powers of the magnitude of 3.0 are rarely obtained in practice, yet some extant procedures of item selection might well reject desirable items that could be so characterized.

The last two charts were included mainly for purposes of illustration. Items with a probability of chance success in excess of .25 would not ordinarily be recommended for use in tailored testing. We have noticed, for example, that the approximation methods become less effective as the value of c increases. Simultaneously the maximum likelihood procedures to obtain estimates of the model parameters become more difficult. Also as c increases, more items are required for validity in the tailoring process.

Discussion

There are certain conditions and assumptions tacit in the effective application of the methods. Since equations and graphs that relate sets of parameters to one another are used, the computed values of the item-test point-biserial correlations (the $\hat{\rho}_{i'\theta}$) and the proportions passing the items (the $\hat{P}_{I'}$) should be based on large samples. The condition can be considered satisfied by a minimum of 2,000 cases. In the computation of the $\hat{\rho}_{i'\theta}$, the total test score is substituted for the latent trait. To do this meaningfully, the items in the test must be homogeneous with respect to the latent ability being measured and sufficient in length to effectively reduce the spurious component of such part-whole correlations. As evidence that

these conditions have been satisfied, the Kuder-Richardson formula 20 reliability coefficient for the test should have a minimum value of .90; and the number of items should be at least 80. It is assumed that the normal ogive mental test model is in reasonable agreement with empirical data. Some evidence has been presented to support this assumption (Lord, 1952; Indow & Samejima, 1966). Further assumed is that the latent trait, underlying ability, is normally distributed. While direct proof of this assumption is elusive, tangential evidence tends to support the reasonableness of the assertion.

If $\hat{\hat{c}}_i$ has been approximated and $\hat{\rho}_{i'\theta}$ and $\hat{P}_{I'}$ have been computed, the appropriate chart may be consulted to obtain the approximate values of $\underline{a}_i$, item discriminatory power, and $\underline{b}_i$, item difficulty. One merely plots the data point for the particular item with respect to $\rho_{i'\theta}$, the ordinate, and $P_{I'}$, the abscissa. The values of $\hat{\hat{\underline{a}}}_i$ and $\hat{\hat{\underline{b}}}_i$ may then be interpolated from the grid system. The total area encompassed by the grid system will enclose the data points corresponding to items of potential value in tailored testing for a fixed value of $\underline{c}$. For instance if $\underline{c}$ is fixed for a set of 100 items, a satisfactory distribution of the $\hat{\hat{\underline{b}}}_i$ would be such that roughly 5 item data points occur between the pairs of grid lines denoting the systematic levels of $\underline{b}$. The $\hat{\hat{b}}_i$ should be rectangularly distributed. For a set of such items, tailored testing is most effective when the overall level of item discriminatory power is highest. Even though a minimum value of .8 has been set for item discriminatory powers tailored testing can be improved in terms of a reduction in the average number of items needed to evaluate examinees at a fixed level of validity when items of higher

discriminatory power are available. The minimum value of .8 assures that tailored testing will maintain an advantage over conventional paper and pencil testing with respect to obtaining a given validity with the fewest possible items.

The original intent of the present paper was to enable the screening of items and to assess of the efficacy of a set of items for tailored testing as well as to provide starting values for maximum likelihood estimation; however, it has been brought to the author's attention that the approximations, the $\hat{\hat{a}}_i$ and $\hat{\hat{b}}_i$, correlate .89 and .97, respectively, with their corresponding maximum likelihood estimates, the $\hat{a}_i$ and $\hat{b}_i$. Since $\hat{\hat{a}}$ was a selection variable, the correlations are quite respectable. Assuming independence of errors of estimation (approximation) in the above coefficients, the correlations between the approximations and true parameters are probably of sufficient magnitude to entirely bypass the expense of maximum likelihood estimation.

The investigator in the study providing the above data was Carl J. Jensema, a doctoral candidate at the University of Washington. The approximations were obtained on 4,950 cases and 98 unscreened mathematics items from the Washington Pre-College Test Battery. The Kuder Richardson formula 20 reliability coefficient for the 98 item test was .93. The maximum likelihood estimates were obtained on 1,000 cases and 58 screened items. The model of analysis was the Birnbaum 3-parameter logistic.

The correlation coefficients between the approximations and the maximum likelihood estimates reflect a strong degree of accord between the corresponding normal ogive and logistic models. In theory, the present normal

ogive approach is more closely related to Bayesian procedures since the distributional form of underlying ability is explicitly stated; whereas, in maximum likelihood procedures, no *a priori* assumption is made regarding the distributional form of the latent trait. Further the correlations would confirm the reasonableness of the fit of the models to the empirical data. Some initial concern in this regard was derived from the speededness of the tests in which the items occurred and the instructions under which the items were given that tended to discourage random guessing, an integral assumption in the models. Moreover, these substantial correlations were obtained despite the serious questioning of the mathematical properties of the maximum likelihood solution in the instance of the Birnbaum 3-parameter model by at least one author (Samejima, 1970). In brief, certain response vectors do not provide unique maximum likelihood estimates. In the study alluded to here, this was most in evidence for response vectors associated with scores in the chance score range. What this amounts to during analyses is the truncation of the lower end of the distribution of latent ability. Since the $\hat{\theta}$ are rescaled to a mean of zero and a standard deviation of one, the truncation would have the predictable effects of reducing the $\hat{a}_i$ and increasing the absolute value of the $\hat{b}_i$ in the negative range. Interestingly enough, despite the magnitude of the correlations, these effects are in evidence in the comparison of the maximum likelihood estimates and the approximations. It might well be that the *heuristic estimates* obtained through the present approximation method are to be preferred to maximum likelihood estimates where distortion of the estimates is artifactually introduced by the nature of the analysis.

A Monte Carlo investigation is currently under way to evaluate the direct use of the heuristic estimates in the Bayesian tailored testing procedures derived by Owen (1969). In effect, tailored testing is to be simulated under conditions (1) where all parameters are known, and (2) where only the item parameters have been approximated. An evaluation of the loss in validity occasioned by moving from condition (1) to condition (2) is seen to be of substantial importance.

References

Birnbaum, A. Part 5. Some latent trait models and their use in inferring an examinee's ability. In F. M. Lord and M. R. Novick, Statistical theories of mental test scores, Reading, Mass.: Addison-Wesley, 1968.

Brogden, H. E. Variation in test validity with variation in the distribution of item difficulties, number of items, and degree of their intercorrelation. Psychometrika, 1946, 11, 197-214.

Brogden, H. E. Latent ability and the structure of tests. Mimeographed. West Lafayette, Ind.: Purdue University, 1971.

Indow, T., & Samejima, F. On the results obtained by the absolute scaling model and the Lord model in the field of intelligence. Yokohama: Psychological Laboratory, Hiyoshi Campus, Keio University, 1966.

Lord, F. M. A theory of test scores. Psychometric Monograph, 1952, No. 7.

Lord, F. M. An analysis of the verbal scholastic aptitude test using Birnbaum's three parameter logistic model. Educational and Psychological Measurement, 1968, 28, 989-1020. (a)

Lord, F. M. Statistical theories of mental test scores. Reading, Mass.: Addison-Wesley, 1968. (b)

Owen, R. J. A Bayesian approach to tailored testing. Research Bulletin 69-92. Princeton, N. J.: Educational Testing Service, 1969.

Samejima, F. Syndrome response categories and information functions. Paper read at Psychometric Society Meeting, Palo Alto, Calif., March 1970.

Urry, V. W. A Monte Carlo investigation of logistic mental test models. Unpublished doctoral dissertation. Purdue University, 1970.

Urry, V. W. Individualized testing by Bayesian estimation. Seattle: Bureau of Testing, University of Washington, 1971. (Duplicated Report)